Package 'JRF'

October 25, 2016

Version 0.1-4

Date 2015-11-17

Title Joint Random Forest (JRF) for the Simultaneous Estimation of Multiple Related Networks

Depends R (>= 3.0.0)

Suggests MASS

Description Simultaneous estimation of multiple related networks.

License GPL (>= 2)

URL https://www.r-project.org

NeedsCompilation yes

Author Francesca Petralia Developer [aut, cre],
Pei Wang Developer [aut],
Zhidong Tu Developer [aut],
Won-min Song Developer [aut],
Adele Cutler Developer [ctb],
Leo Breiman Developer [ctb],
Andy Liaw Developer [ctb],
Matthew Wiener Developer [ctb]

Maintainer Francesca Petralia Developer <francesca.petralia@mssm.edu>

Repository CRAN

Date/Publication 2016-10-25 20:31:57

R topics documented:

JRF *Joint Random Forest for the simultaneous estimation of multiple related networks*

Description

Algorithm for the simultaneous estimation of multiple related networks. Some of the functions utilized are a modified version of functions contained in the R package randomForest (A. Liaw and M. Wiener, 2002).

Usage

```
JRF(X, ntree, mtry,genes.name)
```

Arguments

X	List object containing expression data for each class, `X=list(x_1,x_2, ... )` where `x_j` is a `(p x n_j)` matrix with rows corresponding to genes and columns to samples. Missing values are not allowed.
ntree	numeric value: number of trees.
mtry	numeric value: number of predictors to be sampled at each node.
genes.name	vector containing genes name. The order needs to match the rows of `x_j`.

Value

A matrix with `I` rows and `C + 2` columns where `I` is the total number of gene-gene interactions and `C` is the number of classes. The first two columns contain gene names for each interaction while the remaining columns contain importance scores for different classes.

References

Petralia, F., Song, WM., Tu, Z. and Wang, P., A New Method for Joint Network Analysis Reveals Common and Different Co-Expression Patterns Among Genes and Proteins in Breast Cancer, submitted

A. Liaw and M. Wiener (2002). Classification and Regression by randomForest. *R News* **2**, 18–22.

Examples

```
# --- Derive weighted networks via JRF

nclasses=2               # number of data sets / classes
n1<-n2<-20               # sample size for each data sets
p<-5                    # number of variables (genes)
genes.name<-paste("G",seq(1,p),sep="")   # genes name

  # --- Generate data sets
```

```
data1<-matrix(rnorm(p*n1),p,n1)       # generate data1
data2<-matrix(rnorm(p*n2),p,n1)       # generate data2

  # --- Standardize variables to mean 0 and variance 1

 data1 <- t(apply(data1, 1, function(x) { (x - mean(x)) / sd(x) } ))
 data2 <- t(apply(data2, 1, function(x) { (x - mean(x)) / sd(x) } ))

  # --- Run JRF and obtain importance score of interactions for each class

 out<-JRF(list(data1,data2),mtry=round(sqrt(p-1)),ntree=1000,genes.name)
```

JRF_network *Compute FDR of importance scores and return class-specific networks.*

Description

This function computes FDR of importance scores and returns class-specific networks.

Usage

```
JRF_network(out.jrf,out.perm,TH)
```

Arguments

out.jrf	output object from function JRF.
out.perm	output object from function Run_permutation.
TH	Threshold for FDR.

Value

out list object containing the estimated gene-gene interactions for each class.

References

Petralia, F., Song, WM., Tu, Z. and Wang, P., A New Method for Joint Network Analysis Reveals Common and Different Co-Expression Patterns Among Genes and Proteins in Breast Cancer, submitted

A. Liaw and M. Wiener (2002). Classification and Regression by randomForest. *R News* **2**, 18–22.

Xie, Y., Pan, W. and Khodursky, A.B., 2005. A note on using permutation-based false discovery rate estimates to compare different analysis methods for microarray data. *Bioinformatics*, **21**(23), pp.4280-4288.

Examples

```
# --- Derive weighted networks via JRF

nclasses=2             # number of data sets / classes
n1<-n2<-20             # sample size for each data sets
p<-5                   # number of variables (genes)
genes.name<-paste("G",seq(1,p),sep="")   # genes name
M=5;                   # total number of permutations
fdr=.001;              # fdr threshold

  # --- Generate data sets

data1<-matrix(rnorm(p*n1),p,n1)       # generate data1
data2<-matrix(rnorm(p*n2),p,n1)       # generate data2
data1[1,]<-2*data1[2,]     # variable 1 and 2 interact under class 1

 # --- Standardize variables to mean 0 and variance 1

 data1 <- t(apply(data1, 1, function(x) { (x - mean(x)) / sd(x) } ))
 data2 <- t(apply(data2, 1, function(x) { (x - mean(x)) / sd(x) } ))

  # --- Run JRF and obtain importance score of interactions for each class

 out<-JRF(list(data1,data2),mtry=round(sqrt(p-1)),ntree=1000,genes.name)

out.perm<-Run_permutation(list(data1,data2),mtry=round(sqrt(p-1)),ntree=1000,genes.name,M)

 final.net<-JRF_network(out,out.perm,fdr)
```

`JRF_permutation` *Derive importance scores for permuted data.*

Description

This function computes importance score for one permuted data set. Sample labels of target genes are randomly permuted and JRF is implemented. Resulting importance scores can be used to derive an estimate of FDR.

Usage

```
JRF_permutation(X, ntree, mtry,genes.name,perm)
```

Arguments

`X`	List object containing expression data for each class, `X=list(x_1,x_2, ... )` where `x_j` is a `(p x n_j)` matrix with rows corresponding to genes and columns to samples. Missing values are not allowed.

ntree	numeric value: number of trees.
mtry	numeric value: number of predictors to be sampled at each node.
genes.name	vector containing genes name. The order needs to match the rows of x_j.
perm	integer: seed for permutation.

Value

A matrix with I rows and C columns with I being the number of total interactions and C the number of classes. Element (i,k) corresponds to the importance score for interaction i under class k.

References

Petralia, F., Song, WM., Tu, Z. and Wang, P., A New Method for Joint Network Analysis Reveals Common and Different Co-Expression Patterns Among Genes and Proteins in Breast Cancer, submitted

A. Liaw and M. Wiener (2002). Classification and Regression by randomForest. *R News* **2**, 18–22.

Examples

```
# --- Derive weighted networks via JRF

nclasses=2               # number of data sets / classes
n1<-n2<-20               # sample size for each data sets
p<-5                   # number of variables (genes)
genes.name<-paste("G",seq(1,p),sep="")   # genes name
perm=1;        # set permutation seed

  # --- Generate data sets

data1<-matrix(rnorm(p*n1),p,n1)       # generate data1
data2<-matrix(rnorm(p*n2),p,n1)       # generate data2

  # --- Standardize variables to mean 0 and variance 1

 data1 <- t(apply(data1, 1, function(x) { (x - mean(x)) / sd(x) } ))
 data2 <- t(apply(data2, 1, function(x) { (x - mean(x)) / sd(x) } ))

  # --- Run JRF and obtain importance score of interactions for each class

out<-JRF_permutation(list(data1,data2),mtry=round(sqrt(p-1)),ntree=1000,genes.name,perm)
```

Run_permutation *Derive importance scores for M permuted data sets.*

Description

This function computes importance score for M permuted data sets. Sample labels of target genes are randomly permuted and JRF is implemented. Resulting importance scores can be used to derive an estimate of FDR.

Usage

```
Run_permutation(X, ntree, mtry,genes.name,M)
```

Arguments

X	List object containing expression data for each class, X=list(x_1,x_2, ...) where x_j is a (p x n_j) matrix with rows corresponding to genes and columns to samples. Rows need to be the same across objects, while samples can vary. Missing values are not allowed.
ntree	numeric value: number of trees.
mtry	numeric value: number of predictors to be sampled at each node.
genes.name	vector containing genes name. The order needs to match the rows of x_j.
M	integer: total number of permutations.

Value

A three dimensional matrix (I x M x C) with I being the number of total interactions, M the number of permutations and C the number of classes. Element (i,j,k) corresponds to the importance score for interaction i, permuted data j and class k.

References

Petralia, F., Song, WM., Tu, Z. and Wang, P., A New Method for Joint Network Analysis Reveals Common and Different Co-Expression Patterns Among Genes and Proteins in Breast Cancer, submitted

A. Liaw and M. Wiener (2002). Classification and Regression by randomForest. *R News* **2**, 18–22.

Examples

```
# --- Derive weighted networks via JRF

nclasses=2               # number of data sets / classes
n1<-n2<-20               # sample size for each data sets
p<-5                   # number of variables (genes)
genes.name<-paste("G",seq(1,p),sep="")   # genes name
perm=1;        # set permutation seed
```

```
  # --- Generate data sets

data1<-matrix(rnorm(p*n1),p,n1)       # generate data1
data2<-matrix(rnorm(p*n2),p,n1)       # generate data2
M=5;

  # --- Standardize variables to mean 0 and variance 1

 data1 <- t(apply(data1, 1, function(x) { (x - mean(x)) / sd(x) } ))
 data2 <- t(apply(data2, 1, function(x) { (x - mean(x)) / sd(x) } ))

  # --- Run JRF and obtain importance score of interactions for each class

 out<-Run_permutation(list(data1,data2),mtry=round(sqrt(p-1)),ntree=1000,genes.name,M)
```

Index